Matrix

15^{th} November 2005

Definition 1. Let $A = \{a_{ij}\}$, $B = \{b_{ij}\}$ and $C = \{c_{ij}\}$ be three matrices. Then $C = A + B$ is called the *addition* of the matrices A and B if $c_{ij} = a_{ij} + b_{ij}$ for all i and j.

§

Definition 2. Let $\mathbf{A} = (a_{ij})$ be an $m \times n$ matrix and $\mathbf{B} = (b_{kl})$ an $n \times p$ matrix. Then the product $\mathbf{AB}$ is an $m \times p$ matrix $\mathbf{C} = (c_{il})$ where,

$$c_{il} = \sum_{k=1}^{n} a_{ik} b_{kl}$$

where $1 \leq i \leq m$ and $1 \leq l \leq p$.

§

Definition 3. The expression obtained by eliminating the n variables $x_1, \ldots, x_n$ from n equations,

$$\left.\begin{aligned} a_{11}x_1 + \ldots + a_{1n}x_n &= 0 \\ &\vdots \\ a_{n1}x_1 + \ldots + a_{nn}x_n &= 0 \end{aligned}\right\} \tag{1}$$

is called the *determinant* of this system of equations, Equation 1. The determinant of matrix A denoted by various different notations, for example $\det(A)$, $|A|$, $\sum(\pm a_1 b_2 c_3 \cdots)$, $D(a_1 b_2 c_3 \cdots)$, or $|a_1 b_2 c_3 \cdots|$.

§

Example 1. For a linear system of three variables, Equation 1 can be written as,

$$\left.\begin{aligned} a_1 x + a_2 y + a_3 z &= 0 \\ b_1 x + b_2 y + b_3 z &= 0 \\ c_1 x + c_2 y + c_3 z &= 0 \end{aligned}\right\} \tag{2}$$

Eliminating x, y and z from Equation 2 gives us,

$$a_1 b_2 c_3 - a_1 b_3 c_2 + a_3 b_1 c_2 - a_2 b_1 c_3 + a_2 b_3 c_1 - a_3 b_2 c_1 = 0$$

Definition 4. A *minor* M_{ij} of any matrix A is the determinant of a reduced matrix obtained by omitting the i^{th} row and the j^{th} column of A.

§

Theorem 1. Determinant can be determined by,

$$|\,A\,| = \sum_{i=1}^{k} a_{ij} C_{ij}$$

where C_{ij} is called the *cofactor* of a_{ij}. The cofactor C_{ij} can also be denoted as a^{ij}, and,

$$C_{ij} = (-1)^{i+j} M_{ij}$$

where M_{ij} is a minor of A.

§

Problem 1. Prove Theorem 1,the theorem for finding the determinant of a matrix by Laplacian expansion.

§

Definition 5. Any pairwisely ordered pair in a permutation p is called a *permutation inversion* in p if $i > j$ and $p_i < p_j$.

§

Theorem 2. Determination of the determinant can also be determined by,

$$|\,A\,| = \sum_{\pi} (-1)^{\mathrm{I}(\pi)} \prod_{i=1}^{n} a_{i,\pi(i)}$$

where π is a permutation which ranges over all permutations of $\{1, \ldots, n\}$, and $\mathrm{I}(\pi)$ is called the *inversion number* of π.

§

Problem 2. Prove the theorem for the determination of determinant by permutation, Theorem 2.

§

Theorem 3. Let a be a constant and A an $n \times n$ matrix. Then,

$$\begin{aligned} |\,aA\,| &= a^n\,|\,A\,| \\ |\,-A\,| &= (-1)^n\,|\,A\,| \\ |\,AB\,| &= |\,A\,|\,|\,B\,| \\ |\,I\,| = |\,AA^{-1}\,| &= |\,A\,|\,|\,A^{-1}\,| = 1 \\ |\,A\,| &= \frac{1}{|\,A^{-1}\,|} \end{aligned}$$

§

Problem 3. Prove Theorem 3, the theorem on properties of determinant.

§

Definition 6. A function in two or more variables is said to be *multilinear* if it is linear in each variable separately.

§

Theorem 4. Determinants of matrix are multilinear in rows and columns.

§

Example 2. Consider an 3×3 matrix,

$$A = \begin{vmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ a_7 & a_8 & a_9 \end{vmatrix}$$

What Theorem 4 says about multilinearity of determinants amounts to saying that,

$$|\,A\,| = \begin{vmatrix} a_1 & 0 & 0 \\ a_4 & a_5 & a_6 \\ a_7 & a_8 & a_9 \end{vmatrix} + \begin{vmatrix} 0 & a_2 & 0 \\ a_4 & a_5 & a_6 \\ a_7 & a_8 & a_9 \end{vmatrix} + \begin{vmatrix} 0 & 0 & a_3 \\ a_4 & a_5 & a_6 \\ a_7 & a_8 & a_9 \end{vmatrix}$$

and

$$|\,A\,| = \begin{vmatrix} a_1 & a_2 & a_3 \\ 0 & a_5 & a_6 \\ 0 & a_8 & a_9 \end{vmatrix} + \begin{vmatrix} 0 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ 0 & a_8 & a_9 \end{vmatrix} + \begin{vmatrix} 0 & a_2 & a_3 \\ 0 & a_5 & a_6 \\ a_7 & a_8 & a_9 \end{vmatrix}$$

Problem 4. Prove the theorem on the multilinearity of determinants, Theorem 4.

§

Definition 7. A *conformal mapping* is a transformation that preserves local angle. The terms *function*, *map* and *transformation* are synonyms.

§

Definition 8. A *similarity transformation* is a conformal mapping the transformation matrix of which is,

$$A' \equiv BAB^{-1}$$

Here A and A' are similar matrices.

§

Theorem 5. Similarity transformation does not change the determinant.

Proof. The proof for this is simply,

$$|\,BAB^{-1}\,| = |\,B\,|\,|\,A\,|\,|\,B^{-1}\,| = |\,B\,|\,|\,A\,|\,\frac{1}{|\,B\,|} = |\,A\,|$$

¶

Example 3.

$$\begin{aligned} |\,B^{-1}AB-\lambda I\,| &= |\,B^{-1}AB-B^{-1}\lambda IB\,| \\ &= |\,B^{-1}(A-\lambda I)B\,| \\ &= |\,B^{-1}\,|\,|\,A-\lambda I\,|\,|\,B\,| \\ &= |\,A-\lambda I\,| \end{aligned}$$

Definition 9. Let A be a square, $n\times n$ matrix. Then the trace of A is,

$$\mathrm{Tr}(A)=\sum_{i=1}^{n} a_{ii}$$

§

Definition 10. The *transpose* of a matrix $A=\{a_{ij}\}$ is $A^T=\{a_{ji}\}$.

§

Definition 11. The *complex conjugate* of a matrix $A=\{a_{ij}\}$ is $\bar{A}=\{\bar{a}_{ij}\}$, where $\bar{a}=p-qi$ if $a=p+qi$.

§

Definition 12. Let $\phi(n)$ or $\phi(x)$ be a positive function, and let $f(n)$ or $f(x)$ be any function. Then $f=O(\phi)$ if $|f|<A\phi$ for some constant A and all values of n and x. Here O is called the *big-O* notation which denotes asymptoticity. The notation $f=O(\phi)$ is read, 'f is of order ϕ'.

§

Theorem 6. Some other properties of the determinant are,

$$|\,A\,| = |\,A^T\,|$$
$$|\,\bar{A}\,| = \overline{|\,A\,|}$$
$$|\,I+\epsilon A\,| = 1+\mathrm{Tr}(A)+O(\epsilon^2),\ \text{for } \epsilon \text{ small}$$

§

Example 4. For a square matrix A,

a. switching rows changes the sign of the determinant
b. factoring out scalars from rows and columns leaves the value of the determinant unchanged
c. adding rows and columns together leaves the determinant's value unchanged
d. multiplying a row by a constant c gives the same determinant multiplied by c
e. if a row or a column is zero, then the determinant is zero
f. if any two rows or columns are equal, then the determinant is zero

Problem 5. Prove the properties of determinant given in Theorem 6.

§

Theorem 7. Some properties of matrix trace are,

$$\mathrm{Tr}(A)=\mathrm{Tr}\left(A^T\right)$$
$$\mathrm{Tr}(A+B)=\mathrm{Tr}(A)+\mathrm{Tr}(B)$$
$$\mathrm{Tr}(\alpha A)=\alpha\,\mathrm{Tr}(A)$$

§

Problem 6. Prove that,

$$\left(A^T\right)^{-1}=\left(A^{-1}\right)^T$$

§

Theorem 8.

$$(AB)^T=B^TA^T$$

Proof.

$$\begin{aligned} \left(B^TA^T\right)_{ij} &= \left(b^T\right)_{ik}\left(a^T\right)_{kj} \\ &= b_{ki}a_{jk} \\ &= a_{jk}b_{ki}=(AB)_{ji}=(AB)^T_{ij} \end{aligned}$$

¶

Definition 13. Let A be a square matrix. Then the *inverse* of A, if it exists, is A^{-1} such that,

$$AA^{-1} = I$$

Furthermore, A is said to be *nonsingular* or *invertible* if its inverse exists, otherwise it is said to be *singular.*

§

Example 5. For a 2×2 matrix,

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

the inverse of A is,

$$A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

If A is a 3×3 matrix, then the inverse of A is,

$$A^{-1} = \frac{1}{|A|} \{\det (m_{ij})\}$$

where m_{ij} is a minor of A.

If A is an $n \times n$ matrix, then A^{-1} can be found by numerical methods, for example Gauss-Jordan elimination, Gaussian elimination, and LU decomposition.

Example 6. The *Gaussian elimination* procedure solves the matrix equation $A\mathbf{x} = \mathbf{b}$ by first forming an augmented matrix equation $[A\,\mathbf{b}]$ and then transform this into an upper triangular matrix $\left[\left\{a'_{ij}\right\} \mathbf{b}'\right]$, where a'_{ij} are all zero except when $i \leq j$. Then,

$$x_i = \frac{1}{a'_{ii}} \left(b'_i - \sum_{j=i+1}^{k} a'_{ij} x_j \right)$$

The *Gauss-Jordan elimination* procedure finds matrix inverse by first forming a matrix $[A\,I]$, and then use the Gaussian elimination to transform this matrix into $[I\,B]$. The result matrix B is in fact A^{-1}.
The *LU decomposition* forms from the matrix A a product LU of two matrices, one lower- while the other upper triangular. This gives us three types of equation to solve, namely when $i < j$, $i = j$ and $i > j$, where i and j are the indices of row and respectively column of the matrix product. Then,

$$A\mathbf{x} = (LU)\mathbf{x} = L(U\mathbf{x}) = \mathbf{b}$$

Letting $\mathbf{y} = U\mathbf{x}$ we have $L\mathbf{y} = \mathbf{b}$, and therefore,

$$y_1 = \frac{b_1}{l_{11}}$$

$$y_i = \frac{y}{l_{ii}} \left(b_i - \sum_{j=1}^{i-1} l_{ij} y_j \right)$$

where $i = 2, \ldots, n$. Then since $U\mathbf{x} = \mathbf{y}$,

$$x_n = \frac{y_n}{u_{nn}}$$

$$x_i = \frac{1}{n_{ii}} \left(y_i - \sum_{j=i+1}^{n} u_{ij} x_j \right)$$

where $i = n - 1, \ldots, 1$.

Theorem 9. Let A and B be two square matrices of equal size. Then,

$$(AB)^{-1} = B^{-1}A^{-1}$$

Proof. Let $C = AB$. Then $B = A^{-1}C$ and $A = CB^{-1}$, therefore,

$$C = AB = \left(CB^{-1}\right)\left(A^{-1}C\right) = CB^{-1}A^{-1}C$$

Hence $CB^{-1}A^{-1} = I$, and thus $B^{-1}A^{-1} = (AB)^{-1}$. ¶

Definition 14. The *Einstein's summation* is the simplification of notation by omitting a summation sign, keeping in mind that repeated indices are implicitly summed over, for example $\sum_i a_{ik}a_{ij}$ becomes $a_{ik}a_{ij}$, and $\sum_i a_i a_i$ becomes $a_i a_i$.

§

Definition 15. The multiplication of two matrices $A = \{a_{ij}\}$ and $B = \{b_{ij}\}$ is the matrix $C = AB$ such that $c_{ik} = a_{ij}b_{jk}$.

§

Theorem 10. The matrix multiplication is associative.

Proof.

$$[(ab)c]_{ij} = (ab)_{ik}c_{kj} = (a_{il}b_{lk})\,c_{kj} = a_{il}\,(b_{lk}c_{kj}) = a_{il}(bc)_{lj} = [a(bc)]_{ij}$$

¶

Example 7. From Theorem 10, which shows us the associativity of matrix multiplication, we could write the multiplication of three matrices as $[abc]_{ij}$, which is the same as writing $a_{il}b_{lk}c_{kj}$. And this applies in a similar manner to the multiplication of four or more matrices.

Theorem 11. If A and B are two square and diagonal matrices, then $AB = BA$. But in general matrix multiplication is not commutative.

§

Problem 7. Prove Theorem 11, which is a theorem about non-commutativity of matrix multiplication.

§

Definition 16. A *block matrix* is a matrix which is is made up of small matrices put together, for example,

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

where A, B, C and D are matrices.

§

Theorem 12. Block matrices may be multiplied together in the usual manner, for example,

$$\begin{bmatrix} A_1 & B_1 \\ C_1 & D_1 \end{bmatrix}\begin{bmatrix} A_2 & B_2 \\ C_2 & D_2 \end{bmatrix} = \begin{bmatrix} A_1A_2 + B_1C_2 & A_1B_2 + B_1D_2 \\ C_1A_2 + D_1C_2 & C_1B_2 + D_1D_2 \end{bmatrix}$$

provided that all the products involved are possible.

§

Problem 8. Prove Theorem 12.

§

Definition 17. Let $A = \{a_{ij}\}$ be an $n \times n$ matrix. Then A is called a *diagonal matrix* if $a_{ij} = 0$ when $i \neq j$. Here $1 \leq i, j \leq n$. In other words, a diagonal matrix has its components in the form $a_{ij} = c_i\delta_{ij}$, where c_i is a constant and δ_{ij} is the Kronecker delta,

$$\delta = \begin{cases} 0, \; i \neq j \\ 1, \; i = j \end{cases}$$

§

Theorem 13. A square matrix A can be diagonalised by the transformation $A = PDP^{-1}$, where P is made up of the eigenvectors of A and D is the diagonal matrix desired.

§

Problem 9. Prove Theorem 13, the theorem on matrix diagonalisation.

§

Example 8. Matrix diagonalisation can greatly help reducing the number of parameters in a system of equations. For instance, the systems $A\mathbf{x} = \mathbf{y}$ when diagonalised becomes $PDP^{-1}\mathbf{x} = \mathbf{y}$, that is $D\mathbf{x}' = \mathbf{y}'$, where $\mathbf{x}' = P^{-1}\mathbf{x}$ and $\mathbf{y}' = P^{-1}\mathbf{y}$. In this case, if A is an $n \times n$ matrix, we say that our new system obtained through the process of diagonalisation has canonicalised from $n \times n$ to n parameters.

Definition 18. A *symmetric* matrix is a square matrix A which satisfies $A^T = A$.

§

Example 9. If A is a symmetric matrix, then $A^{-1}A^T = I$.

Definition 19. Let A be a square matrix. Then A is said to be *orthogonal* if $AA^T = I$.

§

Example 10. Definition 19 is the same as saying that $A^{-1} = A^T$.

Theorem 14. A matrix A is symmetric if it can be expressed as $A = QDQ^T$, where Q is an orthogonal matrix and D is a diagonal matrix.

§

Problem 10. Prove Theorem 14, the problem on symmetric matrix.

§

Example 11. Any square matrix A may be decomposed into two terms added together, that is $A_s + A_a$ where A_s is a symmetric matrix and A_a an antisymmetric matrix, called respectively a *symmetric part* and an *antisymmetric part* of A. Furthermore,

$$A_s = \frac{1}{2}\left(A + A^T\right)$$

and,

$$A_a = \frac{1}{2}\left(A - A^T\right)$$